LEVEL II

12

AD A100695

DEPARTMENT OF STATISTICS

The Ohio State University

OSU

COLUMBUS, OHIO

MODERN DESIGN OF SIMULATION EXPERIMENTS*

by

Edward J. Dudewicz

Technical Report No. 238
Department of Statistics
The Ohio State University
Columbus, Ohio 43210
June 1981

* This research was supported by Office of Naval Research
Contract No. N00014-78-C-0543.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(14) TR-238

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
TR-238	AD-A100695	

4. TITLE (and Subtitle) (6)
Modern Design of Simulation Experiments.

5. TYPE OF REPORT & PERIOD COVERED (9)
Technical Report.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s) (10)
Edward J. Dudewicz

8. CONTRACT OR GRANT NUMBER(s) (15)
N00014-78-C-0543

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Statistics
The Ohio State University
Columbus, Ohio 43210

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR 042-403

11. CONTROLLING OFFICE NAME AND ADDRESS (11)
Office of Naval Research
Department of the Navy
Arlington, Virginia 22217

12. REPORT DATE
June 1981

13. NUMBER OF PAGES
ii + 20

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (12) 22a.

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approval for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
To appear in the Digital Simulation Tools Handbook, to be published by McGraw-Hill Book Co.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Key Words and Phrases: Simulation, Experimental Design, One-at-a-time method, Factorial Design, Fractional Factorial Design, Central Composite Design, Confounding, Efficiency, Optimization, Alternative System Evaluation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
Design of simulation experiments has long been neglected by experimenters, who tend to attempt either to run only k experiments to evaluate k factors, or to run 2**k experiments (often infeasible, always costly). In this paper we survey the possibilities and show how to design for valid inferences with practical-sized experiments.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

H0655

MODERN DESIGN OF SIMULATION EXPERIMENTS*

by

Edward J. Dudewicz

Department of Statistics
The Ohio State University
Columbus, Ohio 43210

INTRODUCTION

A "simulation" is an experiment run inside the computer (instead of in what is generally, in the field, termed the "real world") in order to obtain information about a system (existing or proposed), modifications to a system, or several competing systems. Since it is well-known that **much more information can be extracted from experiments** which are carefully designed statistically than can be from experiments which are not statistically designed, statistics has a large role to play in simulation. This chapter deals with that role. This is an especially important chapter for those not used to thinking of their simulations in statistical terms (e.g., those who "run until the money runs out" or who believe that "one run will show me how the system behaves"), as it will enable

---

* This research was supported by Office of Naval Research Contract No. N00014-78-C-0543.

them to obtain valid results where invalid ones were previously obtained, to obtain those results at lower cost than for valid results obtained without a carefully designed experiment, and to obtain valid results for systems where (without using statistical design) the study is computationally infeasible.

As we have said, since simulation data are essentially obtained from an experiment run "inside the computer," we need to know how to design the experiment whereby these data are generated, and we need to know how to analyze the resulting data. Now this comprises a large part of the field commonly termed statistics, to which whole university departments are devoted. Hence, only the most important aspects of design and analysis will be covered here, with a cavaet that, if a simulation study is to be properly designed and executed, in many cases it will be desirable to have associated with the study from the outset a statistician competent in simulation applications.

It is a fact that statistics grew up in a largely agricultural setting where such assumptions as homoscedasticity (equality of variances of observations taken from diverse sources, such as weights of tomotoes from different types of tomato plants) commonly held true, while in simulation these are often violated (e.g., serious heteroscadasticity -- greatly unequal variances -- is more often

the rule in simulation studies, as when a number of diverse alternative job-shop scheduling rules are being simulated). The theory required to handle these new situations has only recently been developed (e.g., for heteroscedasticity [1] and for autocorrelated observations [2]). Thus, many parts of the topics covered here will, for some time to come, be unavailable in almost any texts or books and manuals on statistical methods.

1. Design

1.a. Factorial, fractional-factorial, and one-at-a-time experiments: why and why/not, efficiency.

The pitfalls of lack of statistical design in a simulation experiment include: invalid inferences; valid inferences at substantially increased costs; and inability to complete the study.

For example, suppose a simulation model has been built of output $Y(x_1, x_2)$ of a system which depends on two input variables x_1 (temperature in °F.) and x_2 (reaction time in minutes) [each reader will know similar examples in his/her own field, and may find it helpful to re-cast this example in the terms applicable to one such example]. Suppose that it is standard to operate at $x_1 = 300$, $x_2 = 10$, and the simulation is to evaluate possible gains

of increasing x_1 (by 25°F.) or x_2 (by 2 minutes). Commonly one would run the model at (x_1, x_2) equal to (300, 10), (325, 10), (300, 12) to evaluate the effect of the proposed changes (one-at-a-time method). Suppose one does so, and the results (each based on one run) are as in Table 1. Let us suppose that no variability is

Table 1. Results of one-at-a-time experiment.

(x_1, x_2)	$Y(x_1, x_2)$ from simulation run
(300,10)	100.0
(325, 10)	102.5
(300, 12)	101.0

present (so, untypically, results of one simulation run are typical), that past experience with the system has typically yielded 100.0 output units at (x_1, x_2) = (300, 10), that expensive plant changes are required to implement changes of this magnitude (so we must simulate before making changes in the plant), and that to be economically feasible we must obtain at least a 5% process yield increase from any recommended changes. We estimate the gain from incrementing x_1 to be

$$102.5 - 100.0 = 2.5$$

and the gain from incrementing x_2 to be

$$101.0 - 100.0 = 1.0$$

for a total estimated gain of 2.5 + 1.0 = 3.5 units, less than 5%, so the proposed improvement in process is discarded.

Now the above assumes an output linear in each of x_1 and x_2 (perhaps reasonable in a small range about the usual process operating conditions), but with no synergistic effects (no interaction)... which is often false in the real-world (at least, we should not believe such an assertion without validation, and the design above allows no such verification). For example, an experiment run at (325, 12) might complete our data set to that of Table 2. This set of data has a far different interpretation than

Table 2. Results of 2^2 experiment.

(x_1, x_2)	$Y(x_1, x_2)$ from simulation run
(300, 10)	100.0
(325, 10)	102.5
(300, 12)	101.0
(325, 12)	110.0

does that of Table 1; we now see it will be very profitable (double the needed 5% gain in output) to make the plant modifications needed to run at the higher temperature (x_1 = 325°F) and longer time (x_2 = 12 min.), thus avoiding an invalid inference. The results of Table 2 would be

better displayed as in Table 3, which shows the 2^2 (two factors, each at two levels) design more clearly. Such results as these can be obtained from as simple an under-

Table 3. Results of 2^2 experiment.

$Y(x_1, x_2)$ from simulation run		
x_1 \ x_2	$x_2 = 10$	$x_2 = 12$
$x_1 = 300$	100.0	101.0
$x_1 = 325$	102.5	110.0

lying true function as

$$Y(x_1,x_2) = 100.0+0.1(x_1-300.0)+0.5(x_2-10.0) +0.2(x_1-300.0)(x_2-10.0).$$

As the number n of variables increases, the (invalid) one-at-a-time method will require a number of simulation runs equal to the number of variables; e.g. if in addition to x_1 (temperature) and x_2 (reaction time) we also have variables x_3 (acid concentration) and x_4 (pH) present, then n = 4 simulation runs would be needed by this (invalid) method. In order to assess effects with each variable at the traditional or a higher level, the analog of the experiment in Table 2 would require $2^4 = 16$ simulation runs. As these runs can be very expensive (e.g., in nuclear simulations, where a run can consume several hours of computer time), we now wish to

investigate more sophisticated designs which yield comparably correct inferences at substantially reduced cost.

To perform this task, we must first consider models for yield Y as a function of variables x_1, x_2, x_3, x_4, ... It is reasonable to assume that true mean yield E(Y) may be accurately represented by a polynomial equation of sufficiently high order:

$$
\begin{aligned}
&EY(x_1, x_2, x_3, x_4, \ldots) \\
&= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\
&\quad + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 \\
&\quad + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{44} x_4^2 \\
&\quad + \beta_{123} x_1 x_2 x_3 + \beta_{124} x_1 x_2 x_4 + \beta_{134} x_1 x_3 x_4 + \beta_{234} x_2 x_3 x_4 \\
&\quad + \beta_{1234} x_1 x_2 x_3 x_4 + \ldots .
\end{aligned}
$$

Different choices of the points $(x_1, x_2, x_3, x_4, \ldots)$ at which simulation runs should be made (each such choice is called an experimental design) allow us to estimate various of the β's (singly or in combination), while assuming others are negligible (which can often be tested). Typically one assumes terms of higher than second order are negligible ($0 = \beta_{123} = \beta_{124} = \beta_{134} = \beta_{234} = \beta_{1234} = \ldots$). The coefficients β_1, β_2, β_3, β_4 are called main effects of variables x_1, x_2, x_3, x_4, while β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34} are called 2-factor interactions (2 f.i.)

of the variables involved in the subscripts. An experimental design is called a Resolution III design if no main effect is confounded with (not able to be estimated separately from) any other main effect, but main effects are confounded with 2 f.i. and 2 f.i. with each other. In a Resolution IV design no main effect is confounded with any other main effect or 2 f.i., but 2 f.i.'s are confounded with each other. Thus, Resolution III will allow us to fit a model $\beta_0 + \sum \beta_i x_i$ in factors x_i, but interaction will bias the fit. A resolution IV design will allow a fit unbiased by 2 f.i., i.e. $\hat{\beta}_i$ will estimate β_i (not $\beta_i \pm$ some of the β_{ij}'s of $x_i x_j$ interactions). The one-at-a-time method yields a resolution IV design, and its disadvantages have been noted above.

1.b. Screening experiments.

In screening experiments one attempts to use designs which allow one to find out (at relatively modest cost, i.e. no. of simulation runs) which of the variables x_1, x_2, x_3, x_4, ... are most influential as to one's system's output in very few experiments by using a Resolution III design (of course one then gives up almost all model fitting ability, the intention being to run a more extensive experiment later on the few highly influential

variables identified). For example, one design commonly used is the Resolution III Plackett-Burman design which can study 7 variables x_1, x_2, x_3, x_4, x_5, x_6, x_7 in 8 simulation runs, shown in Table 4. In this design,

Table 4. Resolution III Plackett-Burman design.

		Variable						
		x_1	x_2	x_3	x_4	x_5	x_6	x_7
Experiment	1	+	+	+	-	+	-	-
	2	-	+	+	+	-	+	-
	3	-	-	+	+	+	-	+
	4	+	-	-	+	+	+	-
	5	-	+	-	-	+	+	+
	6	+	-	+	-	-	+	+
	7	+	+	-	+	-	-	+
	8	-	-	-	-	-	-	-

each experimental variable has 2 levels, low(-) and high(+) (e.g., 300(-) and 325(+) for x_1). In experiment 1, variables 1, 2, 3, 5 are set to their high levels, while variables 4, 6, 7 are set to their low levels. If desired later, this design can be augmented to a Resolution IV design by adding 8 more runs (for a total of 16 runs). One can also study 6 factors in a 12 experiment (Webb) experimental design, at Resolution IV. As another example of a screening design, if we have 5 independent variables then a Webb design of Resolution IV is available (see Table 5)

which involves 10 experiments.

Table 5. 5-variable screening design.

		Variable				
		x_1	x_2	x_3	x_4	x_5
Experiment	1	+	-	-	-	-
	2	-	+	-	-	-
	3	-	-	+	-	-
	4	-	-	-	+	-
	5	-	-	-	-	+
	6	-	+	+	+	+
	7	+	-	+	+	+
	8	+	+	-	+	+
	9	+	+	+	-	+
	10	+	+	+	+	-

Note that the first 5 experiments form a Resolution III design, but the 10-run experiment is much more efficient than the 5-run experiment (it yields estimates with variance $\sigma^2/9$ for main effects, vs. σ^2 for the 5-run one-at-a-time experiment... a 9-fold reduction at a price of 5 more runs, and with 2 f.i. elimination; here σ^2 is the variability inherent in each simulation run's outcome). As a final example of a screening design, with 4 variables x_1, x_2, x_3, x_4 a Webb Resolution IV design with 8

experiments is available (Table 6) (which is a "fractional factorial" design, i.e. a fraction of the 2^4 = 16 experiment factorial design with confounding scheme I = ABCD) which allows one to fit β_0, β_1, β_2, β_3, β_4 unbiased by two-factor interactions (though those interactions themselves are not estimable).

Table 6. 5-variable screening design.

		Variable			
		x_1	x_2	x_3	x_4
Experiment	1	-	-	-	-
	2	+	-	-	+
	3	-	+	-	+
	4	+	+	-	-
	5	-	-	+	+
	6	+	-	+	-
	7	-	+	+	-
	8	+	+	+	+

1.c. Central composite designs and full quadratic models.

Assuming screening has been completed we will wish to use a design which allows for assessment of all main effects, all 2 f.i., and (perhaps) all quadratic effects.

We will illustrate with a 4 variable example.

Here one wishes not simply to find the most important factors, but rather to model system output as a function of design settings. Of course the appropriate design to use depends on one's goals and scope of study, as well as on one's budget. One possibility is an 11 experiment Webb design (Table 7). This design allows one to fit β_0, β_1, β_2, β_3, β_4, β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34} assuming no curvature.

Table 7. 4-variable Webb design.*

		Variable			
		x_1	x_2	x_3	x_4
Experiment	1	0	0	0	0
	2	0	0	1	1
	3	0	1	0	1
	4	0	1	1	0
	5	1	0	0	1
	6	1	0	1	0
	7	1	1	0	0
	8	0	1	1	1
	9	1	0	1	1
	10	1	1	0	1
	11	1	1	1	0

* "0" denotes low (-), "1" denotes high (+).

Since curvature cannot often be ruled out a priori, one will usually desire another design (which, however, will require more experiments) unless one's budget is severely constrained. The full factorial 2^4 design (16 experiments) allows one to estimate $\beta_0 + \beta_{11} + \beta_{22} + \beta_{33} + \beta_{44}$, β_1, β_2, β_3, β_4, β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34}, β_{123}, β_{124}, β_{134}, β_{234} and (assuming β_{1234} negligible) have an estimate of experimental error σ^2. [No suitable "fractions" of this design exist, as they all confound one- and/or two-factor effects, so no such easy reduction of number of experiments needed is possible in the 4-factor case.] By adding one experiment, one obtains the 2^4+1 design (17 points), with which one can now separately estimate β_0, $\beta_{11} + \beta_{22} + \beta_{33} + \beta_{44}$, hence (barring cancelling magnitudes) assess the total quadratic effect independently of the response at the center (β_0).

While the above designs are in common use, more recently the Central Composite Design (CCD) has been used to good advantage in such situations. This requires 8 star (or axial) points be added to the 2^4+1 design, for a total of 25 experiments, and allows a full quadratic model to be fitted. Suitable fractions of the 2^4 may be used (i.e. a fraction with at most one 2-factor interaction in any alias set), and (via confounding scheme I = ABC) one can obtain the full quadratic model estimation with

$$\frac{1}{2}2^4 + 1 + 8 = 17 \text{ points.}$$

As an example, suppose we have six variables to consider say x_1, x_2, x_3, x_4, x_5, x_6. A full-factorial approach would require $2^6 = 64$ experiments, which in many cases would not be feasible. However, a CCD can be implemented with only $2^{6-2} + (2)(6) + 1 = 29$ experiments as shown in Table 8. There the levels are "coded" so that "1" means

Table 8. CCD studying 6 variables, in 29 experiments.

Name	Expt. No.	Variable x_1	x_2	x_3	x_4	x_5	x_6
2^{6-2} Points	1	1	-1	-1	1	-1	-1
	2	-1	1	-1	1	-1	-1
	3	-1	-1	1	1	-1	-1
	4	1	1	1	1	-1	-1
	5	1	-1	-1	-1	1	-1
	6	-1	1	-1	-1	1	-1
	7	-1	-1	1	-1	1	-1
	8	1	1	1	-1	1	-1
	9	1	-1	-1	-1	-1	1
	10	-1	1	-1	-1	-1	1
	11	-1	-1	1	-1	-1	1
	12	1	1	1	-1	-1	1
	13	1	-1	-1	1	1	1
	14	-1	1	-1	1	1	1
	15	-1	-1	1	1	1	1
	16	1	1	1	1	1	1
Star	17	$-\alpha$	0	0	0	0	0
	18	α	0	0	0	0	0
	19	0	$-\alpha$	0	0	0	0
	20	0	α	0	0	0	0
	21	0	0	$-\alpha$	0	0	0
	22	0	0	$-\alpha$	0	0	0
	23	0	0	0	$-\alpha$	0	0
	24	0	0	0	α	0	0
	25	0	0	0	0	$-\alpha$	0
	26	0	0	0	0	α	0
	27	0	0	0	0	0	$-\alpha$
	28	0	0	0	0	0	α
Center	29	0	0	0	0	0	0

the "high" level of the variable, "-1" means the "low" level of the variable, and "0" means the average of the high and low levels. "$-\alpha$" and "α" represent multiples of the "low" and "high" levels; e.g., if one takes $\alpha = 1$ (face-centered star points) these are equal to the

respective "low" and "high" levels of the variable in question. Some typical possibilities are given in Table 9.

Table 9. Coding of variable x_1 levels, CCD.

	Level				
	$-\alpha$	1	0	1	α
$\alpha = 1$	300	300	312.5	325	325
$\alpha = 2$	287.5	300	312.5	325	337.5

If one has bounds L_i and U_i on variable x_i and wishes to explore the full space, $\alpha = 1$ is often recommended. In other settings $1 < \alpha \leq 2$ is often used. The specifics vary from setting to setting, with $\alpha = 1.5$ being a reasonable compromise for experimenters who do not have access to a statistical design expert.

1.d. Efficient implementation.

It is important to note that the ordering of experiments in the above tables is not the recommended order in which the simulation runs should be made. Since often the "next" run starts with the end of the "last" run's random number stream, runs should be in random order to prevent systematic effects (from possible deviations from randomness of one's random

number generator) from systematically biasing the results. A carefully tested and chosen random number generator is essential, and an existing generator should not be used without extensive validation (available in [3]).

Note that (unless one has prior knowledge of how his system will use random numbers and correlation, which is rare) use of variance reduction techniques (see Chapter D) will not often be appropriate.

1.e. Two-stage and sequential designs.

The designs explored so far are reasonable ones to use in homoscedastic systems. In heteroscedastic systems (where σ^2 is a function of the levels of x_1, x_2, ...), other designs are called for. Most commonly one finds that if x_1, x_2, ... represent levels of a continuous variable (such as temperature), then the designs given thus far are appropriate even if σ^2 varies slightly as x_1, x_2, ... are changed.

However if x_1, x_2, ... represent the presence (1) or absence (0) of an attribute, then new designs for selecting the best combination of attributes re: system performance are called for. See [4] for some details.

2. Analysis.

The designs presented in section 1 are appropriate for situations where one wants to know "which of k factors are most important in determining my system's output?" or "What model for system output, in k input factors, should be subjected to optimization (Chapter F)?" Design and analysis of the simulation experiment when one's goal is not to answer the above questions, but rather others, such as: a. What is the long-run mean of my system? (which may be answered using a transformation-based analysis, and involves questions of normality and run-in time, and leads us to two-stage and regenerative approaches); b. Which system parameters have significant effects? (which leads us into analysis of variance (ANOVA), and where transformations should generally not be used); c. How different are the various systems' performances? (which requires simultaneous interval estimates); and d. Which is the best system (or set of system parameters)? (which requires the new methodology of ranking-and-selection procedures), is a subject now undergoing rapid development. While such problems have traditionally been approached in the past with transformations, one-stage procedures, or ANOVA, the pitfalls of some of these traditions and of equal-sample-sizes, and the new methods

recently developed which should be used in the future, are explored in [1].

Towards the future, ongoing developments in multivariate analogs of the design and analysis procedures presented should lead to future procedures which are able to simultaneously consider several output characteristics.

REFERENCES

[1] Dudewicz, Edward J., Heteroscedasticity, Encyclopedia of Statistical Sciences, Volume 3 (edited by N. L. Johnson, S. Kotz, and C. B. Read), John Wiley & Sons, Inc., New York, 1982.

[2] Dudewicz, Edward J. and Zaino, N. A., Jr., Allowance for Correlation in Setting Simulation Run-Length via Ranking-and-Selection Procedures, TIMS Studies in the Management Sciences, Volume 7, 1977, pp. 51-61.

[3] Dudewicz, Edward J. and Ralley, Thomas G., The Handbook of Random Number Generation and Testing With TESTRAND Computer Code, American Sciences Press, Columbus, Ohio, 1981.

[4] Dudewicz, Edward J., Ranking (Ordering) and Selection: An Overview of How to Select the Best, Technometrics, Volume 22, 1980, pp. 113-119.